E.g. Roll a die until see 2^{nd} "6", then $r=2$, and

$$P\begin{pmatrix}\text{see exactly 10}\\ \text{non-sixes before}\\ \text{see second 6}\end{pmatrix} = \binom{10+2-1}{1}\left(\tfrac{1}{6}\right)^2\left(\tfrac{5}{6}\right)^{10}$$

$P(X=10)$

$$\cong 0.049$$

Now

$$F(\infty) = \sum_0^\infty f(x) = \sum_{x=0}^\infty \binom{r+x-1}{r-1} p^r q^x$$

Notice that if $g(q) = (1-q)^{-t}$ then

$$g'(q) = t(1-q)^{-t-1}, \dots, g^{(n)}(q) = \frac{(t+n-1)!}{(t-1)!}(1-q)^{-t-n}, \dots$$

so

$$(1-q)^{-t} = \sum_0^\infty \frac{g^{(n)}(0)}{n!} q^n = \sum_0^\infty \binom{t+n-1}{n} q^n$$

Also, $\binom{r+x-1}{r-1} = \binom{r+x-1}{x}$

So

$$F(\infty) = p^r \sum_{x=0}^\infty \binom{r+x-1}{x} q^x$$

$$= p^r(1-q)^{-r} = p^r(1-(1-p))^{-r} = 1$$

Hence $F(x)$ is called the <u>negative binomial distribution</u>.

- If n is large and p is close to zero (or 1), then the binomial distribution can be approximated by the Poisson distribution. (If p isn't close to 0 or 1 then we use the normal distribution as approximator.)

The <u>Poisson distribution</u> is given by

$$S = \{0, 1, 2, \dots\} \text{ and } f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ for } \lambda > 0.$$

Notice that

$$F(\infty) = \sum_0^\infty f(x) = e^{-\lambda}\sum_0^\infty \frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

Definition

Count the # events in a given continuous interval, then an approximate Poisson process with parameter $\lambda > 0$ occurs if

(i) # events in non-overlapping intervals is independent,

(ii) $p\begin{pmatrix}\text{exactly 1 event in} \\ \text{a sufficiently short} \\ \text{interval of length } h\end{pmatrix} \cong \lambda h$,

(iii) $p\begin{pmatrix}\geq 2 \text{ events in a} \\ \text{sufficiently short interval}\end{pmatrix} \cong 0$.

Now suppose we have an approx. Poisson process. Find $p(X=k)$ where k = # events in an interval of 'unit length'. To do this, partition a 'unit interval' into n chunks, then by (ii), for n sufficiently large,

$$p(\text{exactly 1 event in subinterval of length } 1/n) \cong \lambda \tfrac{1}{n},$$

and (iii) implies

$$p(\geq 2 \text{ events in } \tfrac{1}{n} \text{ subinterval}) \cong 0.$$

Now treat each $1/n$ subinterval as a Bernoulli trial, then the binomial distribution gives

$$p(X=k) \cong \binom{n}{k} (\lambda/n)^k (1-\lambda/n)^{n-k}$$

(it's approximate only because the probability $\cong \lambda/n$).

So as $n \to \infty$

$$p(X=k) \xrightarrow{\sim} \lim_{n\to\infty} \frac{n(n-1)\ldots(n-k+1)}{k!} \frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n(n-1)\ldots(n-k+1)}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= 1 \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot 1 = \frac{\lambda^k e^{-\lambda}}{k!}$$

Hence when n is large and λ/n small, we see that the binomial distribution converges to the Poisson.

Eg. Roll a pair of dice $n = 36$ times and record $X = \#$ pairs of 1's.

$$p(\text{pair of 1's in one roll}) = \frac{1}{36} = p$$

So

$$p(X = x, \text{binomial}) = \binom{36}{x}\left(\frac{1}{36}\right)^x\left(1 - \frac{1}{36}\right)^{36-x}$$

$$p(X = x, \text{Poisson}) = \frac{1^x e^{-1}}{x!} \quad \text{with } \lambda = 36 \cdot \frac{1}{36} = 1$$

so that we can expect 1 event in 36 rolls.

To illustrate ...

x	binomial	Poisson
0	0.363	0.368
1	0.373	0.368
2	0.187	0.184
3	0.060	0.061
4	0.014	0.015
5	0.003	0.003
6	0.000	0.001

But note also that $S_{\text{binomial}} = \{0, 1, 2, \ldots, 36\}$ and $S_{\text{Poisson}} = \{0, 1, 2, \ldots\}$.

Eg. Phone calls arrive on average at 2 every 3 minutes. Assuming Poisson, find $p(2 \text{ calls in } 9 \text{ minutes})$ and $p(\geq 5 \text{ calls in } 9 \text{ minutes})$.

Let $X = \#$ calls in 9 minutes — our 'unit interval'. Here we'd expect 6 calls, so $\lambda = 6$. So

$$p(X = 2) = \frac{6^2 e^{-6}}{2!} = 18 e^{-6}$$

and

$$p(X \geq 5) = 1 - p(X \leq 4)$$
$$= 1 - \sum_0^4 \frac{6^x e^{-6}}{x!}$$
$$\simeq 0.715$$

Definitions

Suppose an experiment has $X = (x_1, x_2)$ as random variable with p.d.f. $f(x_1, x_2)$, then the <u>marginal p.d.f.</u> for x_1 is the p.d.f. for x_1 with no regard for x_2. Notating this by $f_1(x_1)$, we get

$$f_1(x_1) = \sum_{\substack{\text{all } x_2 \text{ with} \\ (x_1, x_2) \in S}} f(x_1, x_2) \quad .$$

Then the <u>marginal distribution</u> for x_1 is $F_1(a_1) = \sum_{x_1 \le a_1} f_1(x_1)$.

For trivariate distributions we can define both

$$f_{12}(x_1, x_2) = \sum_{x_3} f(x_1, x_2, x_3) \text{ and } f_1(x_1) = \sum_{x_2}\sum_{x_3} f(x_1, x_2, x_3)$$

as marginal distributions for (x_1, x_2) and x_1 respectively.

<u>E.g.</u> If

$$f(x_1, x_2) = \binom{n}{x_1 \quad x_2 \quad n-x_1-x_2} p_1^{x_1} p_2^{x_2} p_3^{n-x_1-x_2}$$

then

$$f_1(x_1) = \sum_{x_2=0}^{n-x_1} \binom{n}{x_1 \quad x_2 \quad n-x_1-x_2} p_1^{x_1} p_2^{x_2} p_3^{n-x_1-x_2}$$

$$= \binom{n}{x_1} p_1^{x_1} \sum_{x_2=0}^{n-x_1} \binom{n-x_1}{x_2} p_2^{x_2} p_3^{n-x_1-x_2}$$

$$= \binom{n}{x_1} p_1^{x_1} (p_2 + p_3)^{n-x_1}$$

= probability for n repetitions with probability $= p$.

Definition

x_1 and x_2 are <u>statistically independent</u> if

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) \quad \forall (x_1, x_2) \in S.$$

Hence if a given p.d.f. $f(x_1, x_2)$ factors into $\varphi(x_1)\psi(x_2)$ then x_1 and x_2 <u>are</u> stat. indep.. So defining the <u>conditional p.d.f.</u> $f_1(x_1 | x_2) := \frac{f(x_1, x_2)}{f_2(x_2)}$ we have x_1 and x_2 statistically independent iff $f_1(x_1 | x_2) = f_1(x_1)$, or sim. with x_2.

Characteristics of Distributions

Definition

Let X be a discrete random variable with p.d.f. f on S, and $w(x)$ a 'payoff function' on S. Then the expectation of $w(x)$ is

$$E[w(x)] := \sum_{x \in S} w(x) f(x) \quad \text{if the sum exists.}$$

Notice that

(i) for λ constant, $E[\lambda] = \lambda$.
(ii) for λ constant, $E[\lambda . w(x)] = \lambda . E[w(x)]$.
(iii) $E[w(x) + v(x)] = E[w(x)] + E[v(x)]$.

alias, E is a linear operator

(iv) for x, y stat. indep., $E[w(x) . v(y)] = E[w(x)] . E[v(y)]$.

Definitions

If in the above we let $w(x) = x \ \forall x \in S$, then $E[w(x)] = E[x] = \mu$ gives the mean of S. To measure the spread of a distribution we could use $E[|x-\mu|]$, though this is computationally tricky to manipulate, so we define the variance of S by

$$\operatorname{var}(x) = \sigma^2 := E[(x-\mu)^2]$$

RMS, root mean square

and the standard deviation of S by $\sigma := \sqrt{E[(x-\mu)^2]}$.

Notice that
$$\sigma^2 = E[(x-\mu)^2] = E[x^2 - 2x\mu + \mu^2]$$
$$= E[x^2] - 2\mu E[x] + \mu^2 = E[x^2] - (E[x])^2.$$

Also, if $y = ax + b$, then
$$\mu_y = E[ax+b] = a\mu_x + b,$$
and
$$\sigma_y^2 = E[(ax+b-a\mu_x-b)^2] = E[a^2(x-\mu_x)^2] = a^2\sigma_x^2$$
$$\Rightarrow \quad \sigma_y = |a| \sigma_x.$$

Now suppose x and y stat. indep., then for $z = x+y$,

$$\mu_z = E[x+y] = \mu_x + \mu_y \quad \text{(irrespective of independence)}$$

and

$$\begin{aligned}\sigma_z^2 = E[(z-\mu_z)^2] &= E[(x-\mu_x+y-\mu_y)^2] \\ &= E[(x-\mu_x)^2 + (y-\mu_y)^2 + 2(x-\mu_x)(y-\mu_y)] \\ &= \sigma_x^2 + \sigma_y^2 + 2E[x-\mu_x].E[y-\mu_y] = \sigma_x^2+\sigma_y^2 .\end{aligned}$$

Actually, the mean and variance are just two examples of a collection of characteristics of distributions.

Definitions

The k^{th} moment μ'_k of a distribution is given by

$$\mu'_k := E[x^k],$$

so the mean is the 1^{st} moment. It's actually more helpful to consider horizontally normalised shape information, hence we define the k^{th} central moment μ_k of a distribution by

$$\mu_k := E[(x-\mu)^k],$$

so $\mu_1 = 0$ and μ_2 is the variance. There are some colourfully named descriptors, such as

$$\underline{\text{skewness}} := \frac{\mu_3}{\mu_2^{3/2}} \quad \text{and} \quad \underline{\text{kurtosis}} := \frac{\mu_4}{\mu_2^2} .$$

Notice that the ordinary moments determine the central moments ...

$$\mu_2 = E[(x-\mu)^2] = E[x^2] - (E[x])^2 = \mu'_2 - \mu^2,$$

and

$$\begin{aligned}\mu_3 = E[(x-\mu)^3] &= E[x^3] - 3E[x^2]\mu + 3E[x]\mu^2 - \mu^3 \\ &= \mu'_3 - 3\mu'_2\mu + 2\mu^3 .\end{aligned}$$

Exploiting this, there's a useful device called the <u>moment generating function</u> defined by

$$M_x(t) := E[e^{xt}],$$

which when expanded gives ...

$$M_x(t) = E\left[1 + xt + \frac{x^2t^2}{2!} + \dots + \frac{x^nt^n}{n!} + \dots\right]$$

$$= 1 + \mu t + \mu_2' \frac{t^2}{2!} + \dots + \mu_n' \frac{t^n}{n!} + \dots$$

(hence generating moments), and then we can differentiate to access the specific moments ... $\frac{d^k}{dt^k} M_x(t)\Big|_{t=0} = \mu_k'$. A sometimes helpful trick is to let $R(t) := \log_e M_x(t)$, then since $M_x(0) = 1$, we get $R'(0) = M_x'(0) = \mu$ and $R''(0) = M_x''(0) - M_x'(0)^2 = \sigma^2$.

<u>Examples</u>

- Consider the Poisson distribution, so $S = \{0, 1, 2, \dots\}$ and $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$. Hence

  $$\mu = E[x] = \sum_0^\infty x \frac{\lambda^x e^{-\lambda}}{x!} = \lambda e^{-\lambda} \sum_0^\infty \frac{\lambda^x}{x!} = \lambda.$$

  Notice that $E[e^{xt}] = \sum e^{xt} f(x)$, where the sum is taken over all $x \in S$, so here the moment generating function is

  $$M_x(t) = \sum_0^\infty e^{xt} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_0^\infty \frac{(\lambda e^t)^x}{x!} = e^{\lambda(e^t - 1)}.$$

  Hence

  $$R(t) = \lambda(e^t - 1) \Rightarrow R^{(n)}(t) = \lambda e^t \Rightarrow \mu = \lambda \text{ and } \sigma^2 = \lambda.$$

- Consider the binomial distribution, so $S = \{0, 1, \dots, n\}$. Here the moment generating function is

  $$M_x(t) = \sum_0^n e^{xt} \binom{n}{x} p^x q^{n-x} = (pe^t + q)^n.$$

  Differentiating gives $\mu = np$ and $\sigma^2 = npq$. Notice that if $np = \lambda$ and p is small (so $q \cong 1$) then this matches the Poisson values.